

УДК 004.8

Залуцька О.О.

Хмельницький національний університет

МЕТОД АВТОМАТИЗОВАНОГО ОЦІНЮВАННЯ ВІДПОВІДНОСТІ ТОНАЛЬНОСТІ ВІДГУКІВ НА ТОВАРИ В ІНТЕРНЕТ-МАГАЗИНАХ ДО ЇХ КОРИСТУВАЦЬКОЇ ОЦІНКИ З ВИКОРИСТАННЯМ НЕЙРОМЕРЕЖ

У роботі розроблено та програмно реалізовано метод автоматизованого оцінювання відповідності тональності відгуків на товари повсякденною українською мовою в інтернет-магазинах до їх користувацької оцінки з використанням нейромереж. У процесі розробки методу було досліджено ряд ключових аспектів, включаючи створення маркованого набору даних для навчання нейронної мережі, вибір та налаштування нейромережевого класифікатора, а також побудову семантичної моделі мови. Серед нейронних мереж було обрано й використано архітектуру RoBERTa, яка є більш ефективною для аналізу коротких текстових документів. Було підготовлено власний набір даних, який включав 7656 відгуків. Для оцінки продуктивності запропонованої архітектури були використані метрики точності та втрат. Результати показали, що для об'єднаних багатомовних відгуків точність склала 0,92, і функція втрат мала значення 0,29.

In the paper, the method of correspondence automated evaluation of the tonality of product reviews in everyday Ukrainian in online stores to their user evaluation using neural networks was developed and implemented. In the process of developing the method, a number of key aspects were investigated, including the creation of a labeled dataset for training a neural network, the selection and tuning of a neural network classifier, and the construction of a semantic language model. Among the neural networks, the RoBERTa architecture was chosen and used, which is more efficient for the analysis of short text documents. The proprietary data set was prepared, which included 7,656 reviews. Accuracy and loss metrics were used to evaluate the performance of the proposed architecture. The results showed that for the combined multilingual responses, the accuracy was 0.92 and the loss function was 0.29.

Відгуки в інтернеті є невід'ємною складовою сучасного цифрового середовища. Сьогоднішній споживач має доступ до великого обсягу інформації, серед якої значну роль відіграють відгуки інших користувачів щодо товарів, послуг, закладів харчування, готелів тощо. Такі відгуки можна знайти на різних онлайн-платформах, включаючи вебсайти, соціальні мережі, форуми і спеціалізовані сервіси для рецензій. Як свідчать дослідження [1], відгуки в інтернеті суттєво впливають на прийняття рішень споживачами. При виборі товарів чи послуг потенційні покупці часто звертаються до думок інших користувачів, щоб сформуванати уявлення про якість. Позитивні відгуки можуть сприяти ухваленню рішення про покупку, тоді як негативні можуть відштовхнути споживача.

Також дослідження підтверджують, що більшість покупців довіряють онлайн-відгукам так само, як і рекомендаціям від знайомих чи близьких. Водночас компанії використовують відгуки як цінний інструмент для вдосконалення своїх продуктів і послуг. Моніторинг думок клієнтів дозволяє їм впроваджувати необхідні зміни, враховуючи побажання та зауваження споживачів, що сприяє збереженню та підвищенню репутації.

Для покупців важливо уникати можливих ризиків у процесі придбання товарів. Один із дієвих способів перевірки якості продукту перед покупкою – це аналіз відгуків користувачів, які вже мали досвід його використання. Такі рецензії містять корисну інформацію про характеристики товару, його функціональність, рівень обслуговування та доставку. Користувачі діляться своїми враженнями та рекомендаціями, які стають основою для ухвалення рішень. На сторінках інтернет-магазинів часто публікуються відгуки, доповнені оцінками, наприклад, за шкалою у вигляді зірок, як зображено на рисунку 1. Крім того, покупці можуть додавати фото чи відео, щоб наочно продемонструвати свої враження.

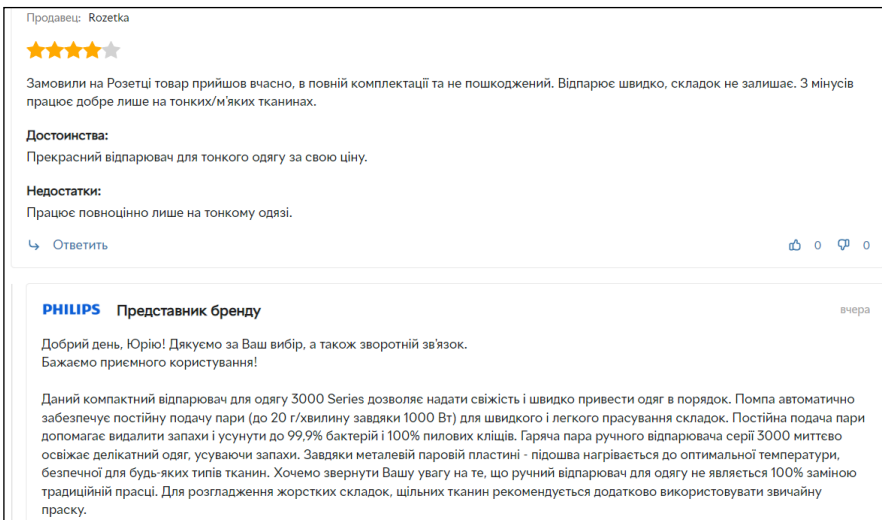


Рисунок 1 – Відгук про товар [1]

Оцінювання товару через відгуки користувачів дозволяє визначити рівень їхньої задоволеності. Однак важливо підходити до цього процесу уважно, аналізуючи різноманітні відгуки, адже вони можуть бути суперечливими. У деяких випадках клієнти можуть залишити відгуки без оцінки через недбалість або випадкову помилку під час виставлення балів. Подібні ситуації можуть призводити до хибного уявлення про якість товару, що може ввести інших покупців в оману і спричинити неправильні рішення при виборі продукту.

Варто пам'ятати, що відгуки й оцінки є суб'єктивними і не завжди об'єктивно відображають якість товару. У відповідь на це багато онлайн-магазинів передбачають можливість редагування оцінок і відгуків через звернення до служби підтримки, якщо вони були залишені помилково [2]..

З огляду на суб'єктивність відгуків, виникає необхідність у впровадженні автоматизованих систем аналізу тональності тексту, які співвідносять зміст відгуку із залишеною оцінкою. Це сприяє створенню більш об'єктивного уявлення про товари та полегшує споживачам процес їхнього вибору.

У сучасній науці особливу увагу приділяють аналізу емоційного забарвлення текстів, що є важливою складовою обробки природної мови. Цей підхід дозволяє оцінювати зміст тексту і розпізнавати настрої, які виражають користувачі. Зростання інтересу до цієї тематики пояснюється широким спектром її застосувань у різних галузях. Аналіз тональності текстів охоплює методи вилучення даних і розпізнавання оцінок, що базуються на використанні алгоритмів машинного навчання.

Нині програми для аналізу тональності текстів стають все популярнішими та знаходять застосування в багатьох сферах. Вони повинні бути здатні точно і надійно визначати емоційний тон тексту, враховуючи реальні почуття авторів. Такі системи мають ефективно працювати з різними мовами, враховувати їхні особливості і забезпечувати обробку значних обсягів інформації.

Як приклад, розглядається програма "AFINN" [3], представлена вперше в жовтні 2011 року. Вона застосовується для аналізу текстів і визначення їхньої емоційної тональності. Приклад роботи цієї програми наведено на рисунку 2.

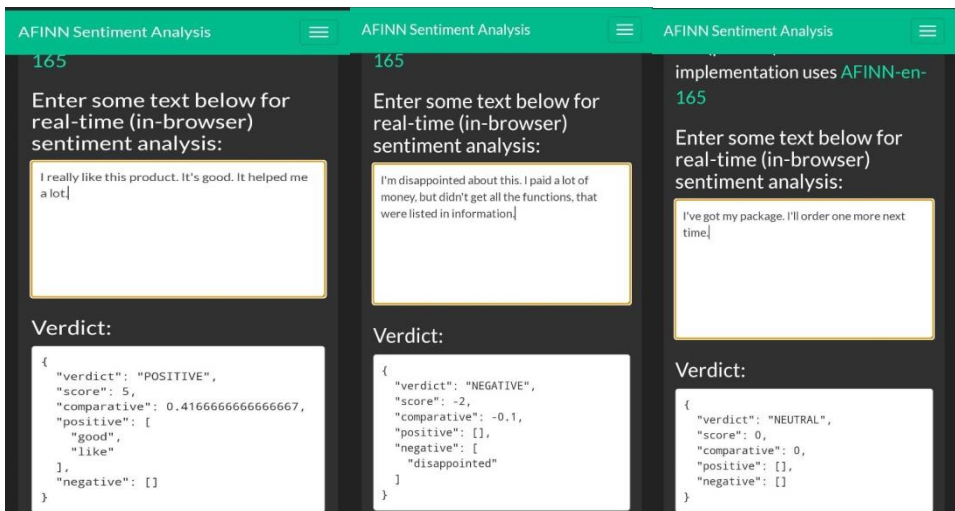


Рисунок 2 – Процес роботи програми [3]

Програма "AFINN" є інструментом для аналізу емоційного забарвлення тексту, що працює на основі словника, де кожному слову англійської мови присвоєно числовий бал. Цей бал визначає емоційний характер слова: позитивні значення вказують на позитивну тональність, негативні – на негативну, а значення, близькі до нуля, свідчать про нейтральність. Принцип роботи програми передбачає обчислення загальної тональності тексту через підсумовування оцінок слів, що у ньому містяться. Чим більша сума балів, тим позитивніший настрій тексту, і навпаки. Серед переваг "AFINN" – доступність словників для англійської мови, можливість доповнення їх новими словами з власними оцінками, простота у використанні та широкодоступність для різних категорій користувачів.

Загалом, "AFINN" є ефективним інструментом для визначення емоційної тональності текстів англійською мовою. Її основа – це числові оцінки слів, що дозволяє оцінити загальний емоційний контекст тексту. Водночас, варто зазначити, що мовна підтримка програми обмежена, оскільки вона розрахована переважно на англійську мову. Крім того, відсутність регулярних оновлень словника може знижувати актуальність цього інструменту в довготривалій перспективі.

Аналіз емоційної тональності відгуків відіграє важливу роль у формуванні розуміння сприйняття клієнтами товарів та послуг. Це дозволяє виявляти тенденції, аналізувати реакцію споживачів, покращувати репутацію бренду і вдосконалювати пропозиції. Такий підхід допомагає знаходити закономірності у відгуках, наприклад, визначати частку позитивних або негативних оцінок, а також найбільш поширені причини задоволення чи незадоволення клієнтів.

Щоб аналізувати відповідність тональності відгуків оцінкам товарів в інтернет-магазинах, застосовуються алгоритми глибокого навчання, побудовані на нейронних мережах штучного інтелекту. Метою таких досліджень є створення систем, що автоматично визначають емоційне забарвлення тексту відгуків і порівнюють його з оцінками, залишеними клієнтами, для оцінки їх відповідності.

Особливістю відгуків у сфері електронної комерції є їх короткий обсяг (зазвичай до 100 слів, а інколи навіть лише 1–3 слова). Багато з них містять неформальні елементи, такі як суржик, жаргон, професіоналізми чи багатомовний контент. Позитивні відгуки часто включають короткі фрази на кшталт "Рекомендую", "Все супер", "Чудовий магазин", тоді як негативні можуть містити вислови типу "Не рекомендую", "Жахливо!".

Серед мовних особливостей відгуків можна виокремити:

- використання жаргону та сленгу (часто у текстах трапляється специфічна лексика, характерна для певних категорій споживачів);
- відхилення від літературної норми (у відгуках часто зустрічаються іншомовні слова, запозичення або перекручені вирази);
- професіоналізми (вживання вузькоспеціалізованих термінів, що потребують розуміння галузевої специфіки);
- згадки про конкретні товари (часто у відгуках присутні назви брендів або продуктів).

Таким чином, аналіз тональності відгуків дозволяє ефективно оцінювати споживчі настрої, що має важливе значення для розвитку бізнесу та вдосконалення його пропозицій.

Урахування цих мовних особливостей також дозволяє покращити якість аналізу та зробити його більш відповідним конкретній галузі, яку представляють відгуки про електронну комерцію. Це може виявитися корисним для підприємств, які бажають краще розуміти відгуки своїх клієнтів і покращити якість обслуговування. Приклад частини відгуку: *"Мені потрібно було вивести USB 3.0 на передню частину корпусу, тому що в мене є флешки USB 3.0, а підходить до задньої частини комп'ютера і вставляти їх незручно, тому що спереду є тільки USB 2.0. Тому я замовив на Rozetka перехідник Chieftec USB 3.0..."*

Багатомовний контент також досить поширений у відгуках. Наприклад, ось відгук, який містить помилки та росіянізми: "Я замовив акумулятор у інтернет-магазині, оскільки перевірів, що у них є хороші відгуки." У цьому реченні виявляються орфографічні помилки, зокрема вони виникають через використання слів, що походять із російської мови. Українська версія цього речення має наступний вигляд: "Я замовив акумулятор у магазині в інтернеті, оскільки переконався, що вони мають високі рейтинги та хороші відгуки."

Ураховуючи ці обмеження, виникає необхідність знайти експериментальні дані, які відповідають вищезазначеним критеріям. Схема виявлення відмінностей між відгуками та користувацькою оцінкою представлена на рисунку 3.

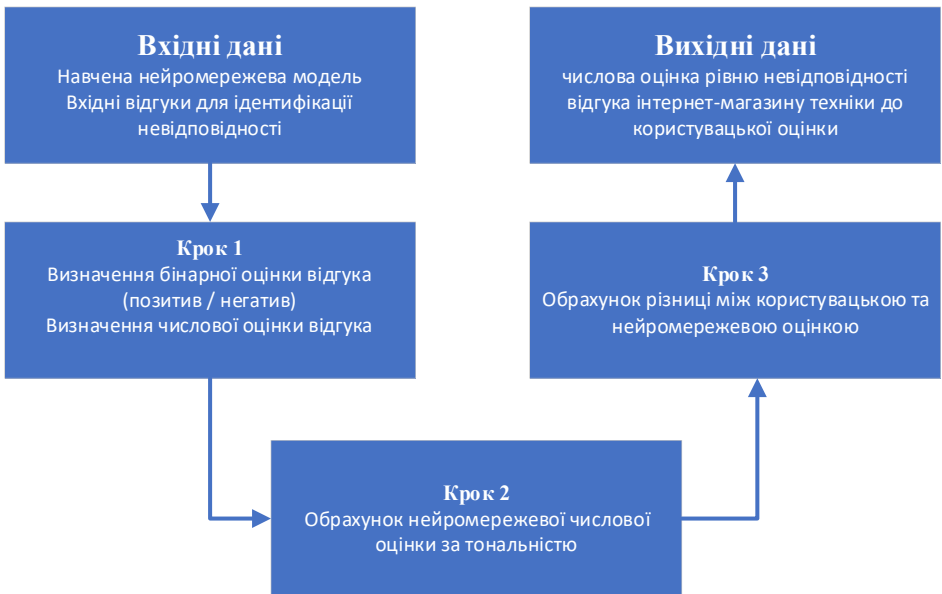


Рисунок 3 – Схема виявлення невідповідності відгука до користувацької оцінки

Початковим етапом методу є визначення бінарної оцінки відгуку, яке включає розділення повідомлень на негативні та позитивні та визначення оцінки відгуку.

Наступним кроком є обрахунок числової оцінки за тональністю за допомогою нейромережі. Для позитивних відгуків числова оцінка обчислюється за допомогою формули:

$$Level = 5 * NeuralNetworkSentimentMark$$

Для негативних відгуків нейромережева оцінка визначається наступним чином:

$$Level = 5 * (1 - NeuralNetworkSentimentMark)$$

На третьому етапі проводиться розрахунок різниці між користувацькою та нейромережевою оцінкою, що виконується за допомогою формули:

$$\Delta = |NeuralNetworkSentimentMark - UserMark|$$

Якщо метод обчислює різницю, абсолютне значення якої перевищує 1.5, відгук вважається невідповідним користувацькій оцінці. Метод дозволяє отримати числову оцінку рівня розбіжності між текстовим описом відгуку клієнта щодо товару в інтернет-магазині техніки та його оцінкою у вигляді числового балу.

На початковому етапі відбувається витяг тексту відгуку з бази даних. Перший крок аналізу передбачає визначення бінарної тональності (позитивна чи негативна) та числової оцінки тексту. Далі розраховується числовий показник емоційного забарвлення відгуку за допомогою нейромережі, як це продемонстровано на рисунку 3. Останній етап полягає у порівнянні користувацької оцінки з оцінкою, яку визначила нейромережева модель. Наприклад, якщо клієнт поставив оцінку 4, а модель вирахувала 1.748, то різниця в абсолютному значенні дорівнює 2.2519. Оскільки це значення перевищує поріг 1.5, відгук класифікується як невідповідний.

Використання глибокої нейромережі дозволяє системі визначити ступінь позитивності або негативності відгуку, що дає можливість користувачам краще зрозуміти емоційний контекст тексту. Це є важливим для споживачів, адже допомагає їм оцінити враження та думки інших клієнтів стосовно товару.

Розроблений метод аналізу розбіжностей між текстовими відгуками та оцінками у категорії технічних товарів в інтернет-магазині базується на використанні глибокої нейромережі. Ця система допомагає виявляти помилки в оцінюванні товарів, аналізуючи текстові описи, та співставляти їх із числовими оцінками, які залишили користувачі. Структуру роботи методу представлено на рисунку 4.

Оснovoю функціонування системи складає мовна модель нейромережі, яка визначає тональність тексту. Для реалізації алгоритму було обрано модель RoBERTa, що є сучасним інструментом для аналізу текстів.

RoBERTa, яке є скороченням від "Robustly optimized BERT approach," є варіацією моделі BERT (Bidirectional Encoder Representations from Transformers), розробленої командою дослідників з Facebook AI [11]. Так само, як BERT, RoBERTa – це мовна модель, яка базується на технології трансформаторів, використовуючи контекст для обробки вхідних текстів та створення контекстуалізованих репрезентацій слів у тексті.

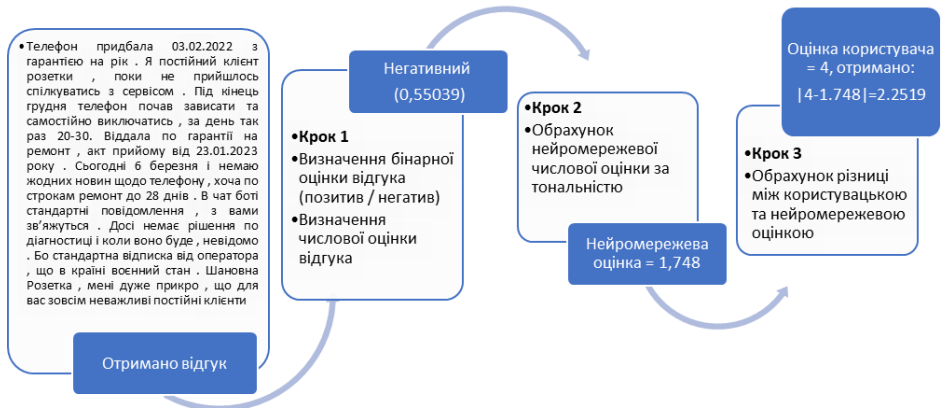


Рисунок 4 – Ілюстрація роботи методу із отриманими результатами

Однією з головних відмінностей RoBERTa від BERT є те, що RoBERTa було навчено на великому обсязі даних і використовувалася більш ефективна процедура навчання. Під час навчання RoBERTa використовує техніку динамічного маскування, яка сприяє покращенню надійних та узагальнених репрезентацій слів. Вибір багатомовних моделей обумовлений тим, що, як вже було зазначено, відгуки можуть містити текст не лише у літературній українській мові.

Конфігурація нейронної мережі на основі обраного набору даних та типу нейронної мережі має структуру, за якої вхідний шар перетворює вхідну текстову інформацію в форму тензору в бібліотеці Keras, іншими словами, у символічний тензороподібний об'єкт, який обгортатиметься атрибутами, що потрібні для побудови моделі Keras на основі вхідних та вихідних даних. Після цього тензор подається на вхід до шару попередньої обробки, включаючи обгортку викликаного об'єкта, для використання як Keras-сумісний шар, побудований на основі задалегідь навченої моделі обробки тексту [4].

Ця модель використовує SentencepieceTokenizer [5, 6], який токенизує тензор строків у кодуванні UTF-8 та діє як інструмент для розділення тексту на токени та їх подальшого об'єднання.

Наступним в шаровій структурі є кодер RoBERTa. Цей шар базується на попередньо навченій моделі "xlm_roberta_multi_cased_L-12_H-768_A-12" [16], яка виникла в результаті масштабного безконтрольного міжмовного навчання (XLM-RoBERTa) [7]. Модель була навчена на 2,5 терабайтів відфільтрованих даних з CommonCrawl, що охоплюють 100 мов [8].

Наступним шаром є шар відсіву, який рандомно обирає певні вхідні одиниці та встановлює їх значення в 0 на кожному навчальному кроці, що допомагає уникнути перенавчання [9]. Вхідні дані, які не дорівнюють 0, масштабуються так, щоб збережено було загальну суму всіх вхідних даних.

Для класифікації настроїв україномовних відгуків про електронну комерцію розглядалися різні підходи, включаючи неймережеві та інші методи. Аналіз наукових публікацій вказує на те, що дослідження, які головним чином використовують словникові інструменти для аналізу настроїв у текстових даних та надають перевагу інтерпретації, мають обмежену точність. Серед неймережевих методів, які були вивчені, BERT-подібні мережі вважаються найбільш ефективними.

BERT був розроблений для того, щоб допомагати комп'ютерам розуміти сенс неоднозначних виразів у тексті, використовуючи контекст для з'ясування того, в якому контексті був написаний текст [10]. Але дослідники виявили, що моделі, такі як ukr-RoBERTa, ukr-ELECTRA та XLM-R large, показують найкращі результати. Зокрема, XLM-R large та ukr-ELECTRA відзначаються в кращому розпізнаванні довших текстів, тоді як ukr-RoBERTa перевершує інші моделі на коротких послідовностях. Оскільки в даному дослідженні аналізуються відгуки з інтернет-платформи "HotLine" [11], які зазвичай є короткими текстовими повідомленнями, то було вирішено використовувати нейронну мережу RoBERTa.

Останнім етапом роботи моделі є класифікація [12, 13], яка здійснюється за допомогою шару Dense. Цей шар формує результат у вигляді значення в діапазоні від 0 до 1, що відображає рівень позитивності відгуку на електронну комерцію, написаного українською мовою [14, 15]. Зокрема, значення 0 означає негативний відгук, а 1 – позитивний. Модель піддається навчанню на сформованій вибірці, використовуючи різні комбінації параметрів, таких як кількість епох, початкове значення Seed і розмір партії (Batch size).

Кількість епох визначає, скільки разів модель оброблятиме навчальні дані. Параметр Seed, встановлений на рівні 42, гарантує стабільність результатів, як це рекомендується в [16], оскільки random_state=42 забезпечує однаковий набір тестових даних при кожному запуску алгоритму. Batch size задає кількість прикладів, які обробляються за одну ітерацію. Через складність заздалегідь визначити оптимальний розмір партії, його підбирають експериментально.

Процес навчання моделі та дослідження проводився за допомогою Python [17]. Після завершення навчання модель зберегли та інтегрували до проекту в

середовищі MS Visual Studio, використовуючи Windows Forms для створення застосунку. Завдяки інтеграції Python та C#, вдалося створити програмний продукт, який автоматизує оцінювання відповідності емоційного забарвлення відгуків на товари до оцінок користувачів, базуючись на технологіях нейронних мереж глибокого навчання.

Для формування навчальної вибірки було використано україномовний контент, який містив марковані дані [19, 20]. В основі оцінювання лежать рейтинги користувачів, де фраза "Не рекомендую" позначає негативний відгук, а "Рекомендую" — позитивний. У вибірці не враховувалися відгуки з іншими оцінками. Для збирання відгуків розробили спеціальне програмне забезпечення на основі бібліотеки Crawler, яке далі обробляли мовою C# і класифікували на дві категорії: "позитивні" та "негативні". Розподіл даних у масиві зображено на рисунку 5.



Рисунок 5 – Діаграми розподілів відгуків у датасеті

Загалом набір даних включає 7656 документів, із яких 6655 документів складають навчальний набір, тоді як для валідації використовувалося 1331 документ (що становить 20% від навчального набору). Особливістю цього набору даних є наявність російських слів, ненормативної лексики та частково російськомовних відгуків. Це пов'язано з тим, що навіть при втраті російською

мовою домінуючої позиції в соціальних мережах після початку війни, вона все ще є присутньою - 37% постів на українській мові порівняно з 63% на російській. Проте варто зауважити, що ця статистика може варіюватися в залежності від конкретних соціальних мереж. Крім того, відгуки часто містять слова з орфографічними помилками.

На рисунку 6 показано діаграму класів, що представляє різні компоненти програмного застосунку.

Програма включатиме в себе форми інтерфейсу користувача, нейромережову модель і набір даних, з яким будуть взаємодіяти різні модулі в межах програмного забезпечення [21].

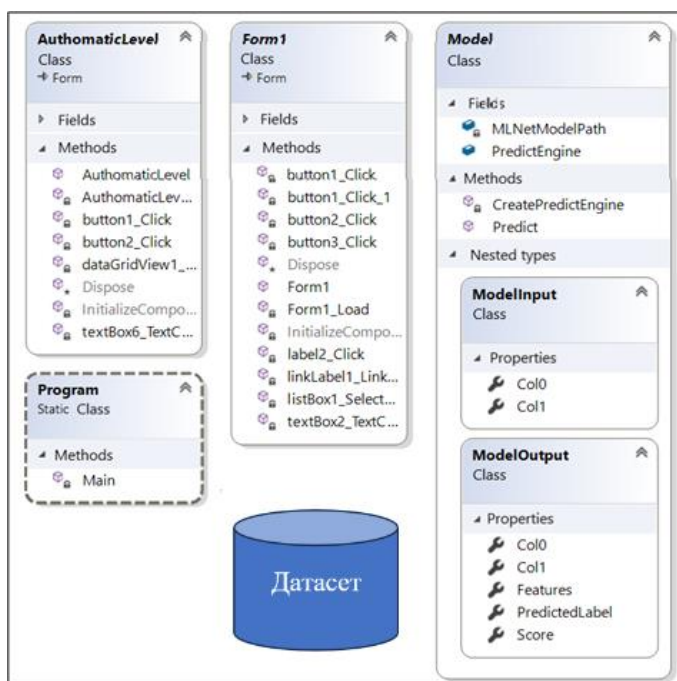


Рисунок 6 – Діаграма класів програмного продукту

Клас "Form1" призначено для спілкування користувача з підсистемою, відповідальною за визначення рівня позитивного відгука користувача. Ця підсистема виконує наступні функції:

- перегляд користувацьких відгуків.
- додавання нових відгуків до бази даних.
- видалення існуючих відгуків із бази даних.

Основна функція цієї форми - визначення емоційного тону текстового відгука та перехід до наступної форми, "Автоматизатор виявлення невідповідності", де автоматизовано виявляються розходження між відгуками та оцінками.

Клас "Автоматизатор виявлення невідповідності до оцінки" виконує наступні завдання:

- вибір відгуку для подальшого аналізу.
- перегляд змісту відгуку та порівняння його емоційного забарвлення з користувацькою оцінкою.
- розрахунок ступеня відповідності відгуку до користувацької оцінки.

Клас "ModelRoberta" містить інформацію про шлях до навченої нейромережевої моделі і відповідає за обчислення, пов'язані із визначенням емоційного забарвлення користувацьких текстових відгуків.

Крім того, в програмному продукті присутній датасет, який містить розмічені відгуки про товари. Цей датасет взаємодіє з усіма модулями системи.

Для реалізації програмного застосунку було створено відповідний інтерфейс. В полі "Текст відгуку" користувач може переглянути вміст кожного текстового повідомлення. В розділі "Обрати відгук" користувач має можливість вибрати конкретний відгук для подальшого дослідження (рисунок 7). У спеціально відведеному полі "Рівень позитиву у відгуку" буде відображатися числова оцінка ступеня емоційного забарвлення відгуку.

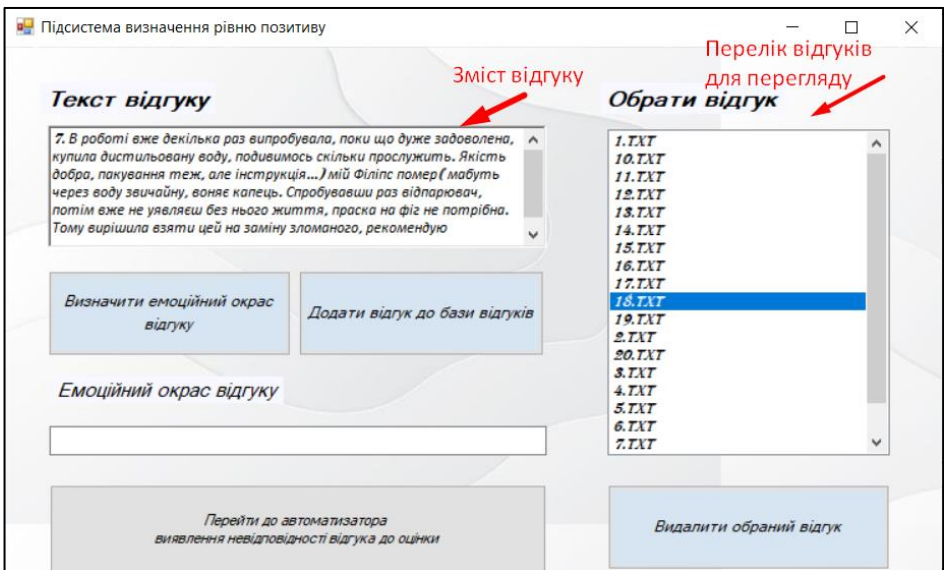


Рисунок 7 – Приклад роботи програми

Після натискання кнопки "Перейти до автоматизатора виявлення невідповідності відгука до оцінки" з'являється нова форма, яка називається "Автоматизатор виявлення невідповідності відгука до оцінки". Тут реалізована можливість перегляду відгука, його емоційного забарвлення та користувацької оцінки в формі таблиці. До цієї форми також додано функціональність, яка дозволяє виконати наступні дії:

- переглянути бінарну користувацьку оцінку відгука;
- переглянути числову оцінку відгука;
- переглянути бінарну нейромережеву оцінку відгука;
- переглянути числову нейромережеву оцінку відгука;
- отримати значення різниці між користувацькою та нейромережевою оцінками.

Після вибору відгука та натискання кнопки "Розрахувати міру відповідності відгука до користувацької оцінки" виконується функція, яка обчислює різницю між нейромережевою оцінкою відгука та оцінкою користувача. Якщо ця різниця перевищує значення 1.5, то відгук вважається невідповідним користувацькій оцінці. Результат роботи функції наведено на рисунку 8.

Автоматизатор виявлення невідповідності відгука до оцінки

Обрати відгуки з бази

Таблиця відгуків користувачів

Відгук	Емоційне забарвлення	Оцінка користувача
Я прихильник техніки Samsung. Хочу написати тут що жодного разу вона мене не розчарувала...	1	5
Пропрацював 3 місяці. Дуже сильно нагрівся і вирубився. Проблема в основній платі, вартість ...	0	1
Користуюсь телефоном 11 місяців. Телефон перегрівався з самого початку, так нічого і не змі...	0	3
Телефон придбала 03.02.2022 з гарантією на рік. Я постійний клієнт розетки, поки не прийшл...	0	4
Купили блендер і він нам сподобався :) Є механізм який запобігає випадковому увімкненню. Р...	1	5
Гарний дизайн, хороші технічні характеристики. При отриманні перевірів - працює, не шумить, ...	1	5
Смартфон слав. Дуже все подобався...	1	5

Розрахувати міру відповідності відгука до користувацької оцінки

Відгук

Телефон придбала 03.02.2022 з гарантією на рік. Я постійний клієнт розетки, поки не прийшлося спілкуватись з сервісом. Під кінець грудня телефон почав зависати та самостійно вимкнутись за день так раз 20-30. Віддала по гарантії на ремонт, акт прийому від 23.01.2023 року. Сьогодні 6 березня і немає жодних новин щодо телефону, хоча по строкам ремонт до 28 днів. В чат боті стандартні повідомлення, з вами зв'язуються. Досі немає рішення по діагностиці і коли воно буде, невідомо. Бо стандартна відписка від оператора, що в країні воєнний стан.

Шановна Розетка, мені дуже прикро, що для вас зовсім неважливі постійні клієнти

Зміст відгука

Користувача оцінка відгука (1 - позитив, 0 - негатив)

Числова користувацька оцінка відгука

Нейромережева оцінка відгука (позитив/негатив)

Числова нейромережева оцінка тональності відгука

Числова нейромережева оцінка відгука

Різниця між користувацькою та нейромережевою оцінками

Відгук не відповідає числовій оцінці користувача

Рисунок 8 – Результат роботи функцій розділу «Автоматизатор виявлення невідповідності відгука до оцінки»

За обраними параметрами були визначені метрики для оцінки продуктивності моделі, включаючи час навчання в секундах, точність і втрати. Для функції втрат використовувалася бінарна крос-ентропійна функція. Точність дослідження визначається як відношення кількості правильних відповідей до загальної кількості відповідей. Отримані результати оцінки продуктивності (час навчання, точність, втрати) різних параметрів налаштування моделі (кількість епох навчання, seed, розмір партії) нейронної мережі наведені в таблиці 1. Експеримент проводився на комп'ютері з процесором Intel Core i7 восьмого покоління, 16 ГБ оперативної пам'яті та відеокартою NVIDIA GeForce MX150.

Таблиця 1 – Параметри перепідготовки класифікатора

Параметри	V1	V2	V3	V4	V5	V6
Кількість епох навчання	3	3	4	5	3	10
Seed	42	42	42	42	42	42
Batch size	64	32	32	32	16	64
Час навчання (sec)	10028	9224	12158	15894	10248	33952
Точність	0.92	0.91	0.91	0.91	0.91	0.92
Loss	0.29	0.31	0.30	0.32	0.31	0.30

Як можна побачити з таблиці 1, модель V1 досягає найвищої точності, яка становить 0,92, та найнижчих втрат, які дорівнюють 0,29. У той час як модель V6 також досягає точності на рівні 0,92, втрати у неї становлять 0,30, при цьому час навчання значно більший.

Навіть при невеликих відхиленнях в точності, майже всі версії навчених моделей дали результати на реальних прикладах, які були подібні до експертних оцінок.

Маючи на увазі, що тестова версія RoBERTa є багатомовною моделлю-трансформером, яка була навчена на двомовних даних, важливо зауважити, що нейронна мережа успішно впоралася з ідентифікацією настрою.

Аналіз відповідей, які не входили в навчальну та тестову вибірки, підтвердив ефективність запропонованої архітектури. Треба відзначити, що навчальна вибірка не була очищена вручну, тому можливі деякі неправильно класифіковані відгуки. Проте це не суттєво вплинуло на загальну точність бінарної класифікації емоційної тональності відгуків, які містять не лише чисту українську мову, але й суржик та двомовні дані.

Результати вказують на те, що використання валідаційної вибірки не призвело до підвищення точності класифікації. Загалом, функція втрат мала

тенденцію незначного зростання після третьої ітерації для валідаційної вибірки. Однак ці результати можуть вказувати на те, що вибірки, ймовірно, не були достатньо обґрунтовано відфільтровані. Проведене тестування нейромережі на відгуках, які відсутні в базі даних, показало майже безпомилкові результати для 40 відгуків, що містять емоції. До позитивної вибірки увійшли такі відгуки, як "Мікрохвильовка", "Купив комп'ютер", "Купив навушники", "Купив пілосос" і так далі. Однак схожі відгуки можуть зустрічатися і в негативній вибірці.

Крім того, аналіз оцінки тональності показав, що нейромережа успішно впоралася із завданням на 40 фразах, яких не було в навчальній або тестовій вибірках. Ці фрази були попередньо оцінені експертом і містили як стилістичні, так і орфографічні помилки, а також були представлені багатомовними даними.

Дослідження вказують на те, що нейромережа дійсно може "розуміти" контекст. В деяких випадках в нейромережі можливі незначні коливання, особливо при нейтральних відгуках, наприклад: "Ціна правильна, наявність теж." Цей відгук був оцінений як "Рекомендую," але нейромережа визначила його як позитивний з майже маргінальною оцінкою 0,505790.

Отже, у роботі було розроблено та програмно реалізовано метод автоматизованого оцінювання відповідності тональності відгуків на товари повсякденною українською мовою в інтернет-магазинах до їх користувацької оцінки з використанням нейромереж глибокого навчання. У процесі розробки методу було досліджено ряд ключових аспектів, включаючи створення маркованого набору даних для навчання нейронної мережі, вибір та налаштування нейромережевого класифікатора, а також побудову семантичної моделі мови. Серед нейронних мереж було обрано й використано архітектуру BERT, яка вважалася однією з найточніших, зокрема її модифікацію RoBERTa, яка є більш ефективною для аналізу коротких текстових документів.

Оскільки україномовні відгуки електронної комерції мають свої унікальні особливості, було підготовлено власний набір даних, який включав 7656 відгуків. Цей набір даних був поділений на навчальну та тестову вибірки, і в кожній з них були представлені як негативні, так і позитивні відгуки. Для оцінки продуктивності запропонованої архітектури були використані метрики точності та втрат. Результати показали, що для об'єднаних багатомовних відгуків точність склала 0,92, і функція втрат мала значення 0,29.

Запропонований підхід має свої обмеження. Він ефективний для коротких текстових рецензій, не довших за 500 слів, написаних українською мовою, включаючи суржик та мовні вкраплення. Майбутні дослідження спрямовані на використання цього класифікатора для оцінювання роботи менеджерів інтернет-магазинів у спілкуванні з клієнтами, впровадження моделей зворотного маркетингу та підвищення ефективності багатомовних класифікаторів.

Перелік посилань

1. Недоліки інтернет-магазинів. URL: <https://www.tvivsvit.com.ua/pokupky-v-interneti-perevahy-ta-nedoliky/>
2. Wei Li, Wei Shao, Shaoxiong Ji, Erik Cambria. "BiERU: Bidirectional emotional recurrent unit for conversational sentiment analysis". *Neurocomputing* (2022): 73-82. doi: 10.1016/j.neucom.2021.09.057.
3. AFINN. URL: <https://darenr.github.io/afinn/>
https://issuu.com/arctionltdchart/docs/what_is_windows_forms_and_its_benefits
4. Tfhub.Dev., Unsupervised Cross-lingual Representation Learning at Scale. xlm_roberta_multi_cased_L-12_H-768_A-12. URL: https://tfhub.dev/jeongukjae/xlm_roberta_multi_cased_L-12_H-768_A-12/1.
5. Tensorflow.Org., Text.SentencepieceTokenizer. URL: https://www.tensorflow.org/text/api_docs/python/text/SentencepieceTokenizer.
6. Tfhub.Dev., Unsupervised Cross-lingual Representation Learning at Scale. xlm_roberta_multi_cased_L-12_H-768_A-12. URL: https://tfhub.dev/jeongukjae/xlm_roberta_multi_cased_L-12_H-768_A-12/1.
7. Huggingface.Co., XLM-RoBERTa (base-sized model). URL: <https://huggingface.co/xlm-roberta-base>
8. ATensorflow.Org., Tf.keras.layers.Dropout. URL: https://www.tensorflow.org/api_docs/python/tf/keras/layers/ Dropout.
9. Sentiment Analysis of Movie Reviews with Google's BERT. URL: <https://medium.com/mllearning-ai/sentiment-analysis-of-movie-reviews-with-googles-bert-c2b97f4217f>
10. Hotline.Ua., Reviews of the store Rozetka. URL: <https://hotline.ua/ua/yp/2476/reviews>
11. Thecleverprogrammer.com., Why Random_state=42 in Machine Learning? URL: https://thecleverprogrammer.com/2020/12/17/why-random_state42-in-machine-learning.
12. Blazhuk V., Mazurets O., Zalutka O. An Approach to Using the mBERT Deep Learning Neural Network Model for Identifying Emotional Components and Communication Intentions. The Impact of Scientific Research on the Development of the Modern World. Proceedings of the XLIV International scientific and practical conference. October 23-25, 2024. Dubrovnik, Croatia. 2024. Pp. 79-84.
13. Mazurets O.V., Sobko O.V., Molchanova M.O., Zalutka O.O., Yurchak A.V. Practical Implementation of Neural Network Method for Stress Features Detection by Social Internet Networks Posts. Global Science: Prospects and Innovations. Proceedings of the II International Scientific and Theoretical Conference «Scientific Review of the Actual Events, Achievements and Problems». May 31, 2024. Berlin, Federal Republic of Germany: International Center of Scientific Research. 2024. Pp. 160-167.
14. Молчанова М.О., Залуцька О.О., Бармак О.В. Метод інтелектуального аналізу тональності текстів. Матеріали XII Всеукраїнської науково-практичної конференції «Глушковські читання». Київ – 2023. с. 113-116.
15. Mazurets O., Zalutka O., Tyschenko O., Bohdanova A. An Approach to Using MobileNet CNN-model for Gesture Recognition. Proceedings of XXIII International Scientific and Practical Conference «Problems of Science and Technology: the Search for Innovative Solutions». May 15-17, 2024. Munich, Germany. 2024. Pp. 59-64.

16. Microsoft ML. URL: <https://learn.microsoft.com/en-us/dotnet/machine-learning/how-does-ml-dotnet-work>
17. Залуцька О.О., Мазурець О.В. Інформаційний портрет ключових термінів у цифрових навчальних матеріалах. Матеріали III Міжнародної науково-практичної конференції «Сучасні інформаційні технології та інноваційні методики навчання: досвід, тенденції, перспективи». Тернопіль, 2019. С.120-122.
18. Войчишин О.О., Залуцька О.О., Попов Ю.М., Купрійчук В.О. Інформаційна технологія автоматизованого формування семантичного ядра цифрових текстів. Збірник наукових праць за матеріалами XIII Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2021». Хмельницький, 2021. с. 298-305.
19. Залуцька О.О., Молчанова М.О., Віт Р.В., Мазурець О.В. Конфігурування нейронної мережі для класифікації емоційної тональності текстової інформації за показниками семантичної зв'язності. Збірник наукових праць за матеріалами XV Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2023». Хмельницький, 2023. с. 102-107.
20. Залуцька О.О., Молчанова М.О., Мазурець О.В., Мельник О.І., Скрипник Т.К. Метод інтелектуального аналізу емоційної тональності текстової інформації для визначення поведінкових намірів нейромережевими засобами. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2023. №5 (325). Т.1. С. 67-73.
21. Krak I., Zalutska O., Molchanova M., Mazurets O., Bahrii R., Sobko O., Barmak O. Abusive Speech Detection Method for Ukrainian Language Used Recurrent Neural Network. CEUR Workshop Proceedings, 2024, vol. 3688, pp. 16-28.